Refine Search

Search Results -

Terms	Documents
(L4 or L5) and (estimat$4 near cardinal$5)	3

Database:
US Pre-Grant Publication Full-Text Database
US Patents Full-Text Database
US OCR Full-Text Database
EPO Abstracts Database
JPO Abstracts Database
Derwent World Patents Index
IBM Technical Disclosure Bulletins

Search: L6

Refine Search

Recall Text   Clear   Interrupt

Search History

DATE: Friday, August 18, 2006 Printable Copy   Create Case

Set Name side by side	Query	Hit Count	Set Name result set
	DB=PGPB,USPT,USOC,EPAB,JPAB,DWPI,TDBD; PLUR=YES; OP=OR		
L6	(L4 or L5) and (estimat$4 near cardinal$5)	3	L6
L5	L1 and L3	125	L5
L4	L1 and L2	129	L4
L3	"data mining"	7454	L3
L2	data near mining	7582	L2
L1	sampl$3 same statistic$4 same collect$4	4328	L1

END OF SEARCH HISTORY

Refine Search

Search Results -

Terms	Documents
L6 and (query$3 near condition)	1

Database:
US Pre-Grant Publication Full-Text Database
US Patents Full-Text Database
US OCR Full-Text Database
EPO Abstracts Database
JPO Abstracts Database
Derwent World Patents Index
IBM Technical Disclosure Bulletins

Search: L7

Refine Search

Recall Text Clear Interrupt

Search History

DATE: Friday, August 18, 2006   Printable Copy   Create Case

Set Name side by side	Query	Hit Count	Set Name result set
	DB=PGPB,USPT,USOC,EPAB,JPAB,DWPI,TDBD; PLUR=YES; OP=OR		
L7	L6 and (query$3 near condition)	1	L7
L6	(L4 or L5) and (estimat$4 near cardinal$5)	3	L6
L5	L1 and L3	125	L5
L4	L1 and L2	129	L4
L3	"data mining"	7454	L3
L2	data near mining	7582	L2
L1	sampl$3 same statistic$4 same collect$4	4328	L1

END OF SEARCH HISTORY

Refine Search

Search Results -

Terms	Documents
L14 and (regression near function)	1

Database:
```
US Pre-Grant Publication Full-Text Database
US Patents Full-Text Database
US OCR Full-Text Database
EPO Abstracts Database
JPO Abstracts Database
Derwent World Patents Index
IBM Technical Disclosure Bulletins
```

Search: L15

[Refine Search]

[Recall Text ⬍]   [Clear]   [Interrupt]

Search History

DATE: Friday, August 18, 2006   Printable Copy   Create Case

Set Name side by side	Query	Hit Count	Set Name result set
DB=PGPB,USPT,USOC,EPAB,JPAB,DWPI,TDBD; PLUR=YES; OP=OR			
L15	L14 and (regression near function)	1	L15
L14	L13 and (calculat$4 same estimat$4 same query$3)	2	L14
L13	L10 and (estimat$4 same query$3)	7	L13
L12	L11 not L6	3	L12
L11	L10 and (query$3 near condition)	4	L11
L10	L9 and tun$3	40	L10
L9	L8 and (("data set") or dataset or (data near set))	230	L9
L8	sampl$3 and queries and cardinal$4 and min$3	436	L8
L7	L6 and (query$3 near condition)	1	L7
L6	(L4 or L5) and (estimat$4 near cardinal$5)	3	L6
L5	L1 and L3	125	L5
L4	L1 and L2	129	L4
L3	"data mining"	7454	L3
L2	data near mining	7582	L2

L1   sampl$3 same statistic$4 same collect$4   4328   L1

END OF SEARCH HISTORY

Hit List

Clear   Generate Collection   Print   Fwd Refs   Bkwd Refs

Generate OACS

Search Results - Record(s) 1 through 1 of 1 returned.

☐ 1. Document ID: US 20040128287 A1

L15: Entry 1 of 1   File: PGPB   .   Jul 1, 2004

PGPUB-DOCUMENT-NUMBER: 20040128287
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20040128287 A1

TITLE: Self tuning database retrieval optimization using regression functions

| Full | Title | Citation | Front | Review | Classification | Date | Reference | Sequences | Attachments | Claims | KWIC | Draw Dr |

Clear   Generate Collection   Print   Fwd Refs   Bkwd Refs   Generate OACS

Terms	Documents
L14 and (regression near function) .	. 1

Display Format: [-] Change Format

Refine Search

Search Results -

Terms	Documents
L24 not L19	0

Database:
US Pre-Grant Publication Full-Text Database
US Patents Full-Text Database
US OCR Full-Text Database
EPO Abstracts Database
JPO Abstracts Database
Derwent World Patents Index
IBM Technical Disclosure Bulletins

Search: L25

Refine Search

Recall Text   Clear   Interrupt

Search History

DATE: Friday, August 18, 2006   Printable Copy   Create Case

Set Name side by side	Query	Hit Count	Set Name result set
	DB=PGPB,USPT,USOC,EPAB,JPAB,DWPI,TDBD; PLUR=YES; OP=OR		
L25	L24 not L19	0	L25
L24	L23 and L17	2	L24
L23	(L20 or L21 or L22) and L16	193	L23
L22	707/200-206.ccls.	7569	L22
L21	707/100-104.1.ccls.	14940	L21
L20	707/1-10.ccls.	22115	L20
L19	L16 and L17	2	L19
L18	L1 and L16 and L17	1	L18
L17	regression near function	680	L17
L16	calculat$4 same estimat$4 same query$3	509	L16
L15	L14 and (regression near function)	1	L15
L14	L13 and (calculat$4 same estimat$4 same query$3)	2	L14
L13	L10 and (estimat$4 same query$3)	7	L13
L12	L11 not L6	3	L12

L11	L10 and (query$3 near condition)	4	L11
L10	L9 and tun$3	40	L10
L9	L8 and (("data set") or dataset or (data near set))	230	L9
L8	sampl$3 and queries and cardinal$4 and min$3	436	L8
L7	L6 and (query$3 near condition)	1	L7
L6	(L4 or L5) and (estimat$4 near cardinal$5)	3	L6
L5	L1 and L3	125	L5
L4	L1 and L2	129	L4
L3	"data mining"	7454	L3
L2	data near mining	7582	L2
L1	sampl$3 same statistic$4 same collect$4	4328	L1

END OF SEARCH HISTORY

Refine Search

Search Results -

Terms	Documents
L7 and L1	0

Database:
US Pre-Grant Publication Full-Text Database
US Patents Full-Text Database
US OCR Full-Text Database
EPO Abstracts Database
JPO Abstracts Database
Derwent World Patents Index
IBM Technical Disclosure Bulletins

Search: L10

Refine Search

Recall Text   Clear   Interrupt

Search History

DATE: Friday, August 18, 2006   Printable Copy   Create Case

Set Name side by side	Query	Hit Count	Set Name result set
	DB=PGPB,USPT,USOC,EPAB,JPAB,DWPI,TDBD; PLUR=YES; OP=OR		
L10	L7 and L1	0	L10
L9	L6 and L1	0	L9
L8	(L4 or L5) and L1	0	L8
L7	708/7.ccls.	169	L7
L6	708/6.ccls.	89	L6
L5	708/2-3.ccls.	276	L5
L4	715/538.ccls.	63	L4
L3	L1 and L2	2	L3
L2	regression near function	680	L2
L1	calculat$4 same estimat$4 same query$3	509	L1

END OF SEARCH HISTORY

Refine Search

Search Results -

Terms	Documents
L16 and (min$3 with model)	1

Database:
```
US Pre-Grant Publication Full-Text Database
US Patents Full-Text Database
US OCR Full-Text Database
EPO Abstracts Database
JPO Abstracts Database
Derwent World Patents Index
IBM Technical Disclosure Bulletins
```

Search: L17

[Refine Search] [Recall Text] [Clear] [Interrupt]

Search History

DATE: Friday, August 18, 2006   Printable Copy   Create Case

Set Name	Query	Hit Count	Set Name
side by side			result set
	DB=PGPB,USPT,USOC,EPAB,JPAB,DWPI,TDBD; PLUR=YES; OP=OR		
L17	L16 and (min$3 with model)	1	L17
L16	(L12 or L13) and ((frequency or frequencies) near occurrence)	5	L16
L15	L14 and (regression near function)	1	L15
L14	(L12 or L13) and regression	4	L14
L13	L12 not L11	5	L13
L12	L1 and (sampl$3 same querie$1)	65	L12
L11	L1 and (sampl$ same querie$1)	60	L11
L10	L7 and L1	0	L10
L9	L6 and L1	0	L9
L8	(L4 or L5) and L1	0	L8
L7	708/7.ccls.	169	L7
L6	708/6.ccls.	89	L6
L5	708/2-3.ccls.	276	L5
L4	715/538.ccls.	63	L4

L3	L1 and L2	2	L3
L2	regression near function	680	L2
L1	calculat$4 same estimat$4 same query$3	509	L1

END OF SEARCH HISTORY

Hit List

Clear   Generate Collection   Print   Fwd Refs   Bkwd Refs
Generate OACS

Search Results - Record(s) 1 through 1 of 1 returned.

☐ 1. Document ID: US 20040128287 A1

```
L17: Entry 1 of 1                    File: PGPB              Jul 1, 2004

PGPUB-DOCUMENT-NUMBER: 20040128287
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20040128287 A1

TITLE: Self tuning database retrieval optimization using regression functions
```

| Full | Title | Citation | Front | Review | Classification | Date | Reference | Sequences | Attachments | Claims | KWIC | Draw D |

Clear   Generate Collection   Print   Fwd Refs   Bkwd Refs   Generate OACS

Terms	Documents
L16 and (min$3 with model)	1

Display Format: [-]   Change Format

Refine Search

Search Results -

Terms	Documents
L4 and (mining near model)	1

Database:

US Pre-Grant Publication Full-Text Database
US Patents Full-Text Database
US OCR Full-Text Database
EPO Abstracts Database
JPO Abstracts Database
Derwent World Patents Index
IBM Technical Disclosure Bulletins

Search:

L5

[Refine Search]

[Recall Text] [Clear] [Interrupt]

Search History

DATE: Friday, August 18, 2006   Printable Copy   Create Case

Set Name side by side	Query	Hit Count	Set Name result set
	DB=PGPB,USPT,USOC,EPAB,JPAB,DWPI,TDBD; PLUR=YES; OP=OR		
L5	L4 and (mining near model)	1	L5
L4	L3 and (estimat$4 near (cardinality or cardinalities))	20	L4
L3	L2 and (estimat$4 with (cardinality or cardinalities))	27	L3
L2	L1 and model	273	L2
L1	calculat$4 same estimat$4 same query$3	509	L1

END OF SEARCH HISTORY

Hit List

Clear   Generate Collection   Print   Fwd Refs   Bkwd Refs

Generate OACS

Search Results - Record(s) 1 through 1 of 1 returned.

☐ 1. Document ID: US 20040128287 A1

L5: Entry 1 of 1   File: PGPB   Jul 1, 2004

PGPUB-DOCUMENT-NUMBER: 20040128287
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20040128287 A1

TITLE: Self tuning database retrieval optimization using regression functions

Full | Title | Citation | Front | Review | Classification | Date | Reference | Sequences | Attachments | Claims | KWIC | Draw De

Clear   Generate Collection   Print   Fwd Refs   Bkwd Refs   Generate OACS

Terms	Documents
L4 and (mining near model)	1

Display Format: |- | Change Format

Refine Search

Search Results -

Terms	Documents
L9 and (regression same function same (updat$3 or chang$3 or modif$4))	2

Database:
```
US Pre-Grant Publication Full-Text Database
US Patents Full-Text Database
US OCR Full-Text Database
EPO Abstracts Database
JPO Abstracts Database
Derwent World Patents Index
IBM Technical Disclosure Bulletins
```

Search: L12

[Refine Search]

[Recall Text] [Clear] [Interrupt]

Search History

DATE: Friday, August 18, 2006   Printable Copy   Create Case

Set Name side by side	Query	Hit Count	Set Name result set
DB=PGPB,USPT,USOC,EPAB,JPAB,DWPI,TDBD; PLUR=YES; OP=OR			
L12	L9 and (regression same function same (updat$3 or chang$3 or modif$4))	2	L12
L11	L9 and (regression same function same (updat$3 or chang$3 or modif$8))	2	L11
L10	L9 and (regression same function same (updat$3 or chang$3 or modif$))	2	L10
L9	L8 and L4	19	L9
L8	707/$.ccls.	37005	L8
L7	L6 and L4	0	L7
L6	708/$.ccls.	28072	L6
L5	L4 and (mining near model)	1	L5
L4	L3 and (estimat$4 near (cardinality or cardinalities))	20	L4
L3	L2 and (estimat$4 with (cardinality or cardinalities))	27	L3

L2	L1 and model	273	L2
L1	calculat$4 same estimat$4 same query$3	509	L1

END OF SEARCH HISTORY

Hit List

Clear   Generate Collection   Print   Fwd Refs   Bkwd Refs
Generate OACS

Search Results - Record(s) 1 through 2 of 2 returned.

☐ 1.   Document ID: US 20050097078 A1
. L12: Entry 1 of 2   File: PGPB   May 5, 2005

PGPUB-DOCUMENT-NUMBER: 20050097078
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20050097078 A1

TITLE: System, method, and computer program product for progressive query
processing

| Full | Title | Citation | Front | Review | Classification | Date | Reference | Sequences | Attachments | Claims | KWIC | Draw D |

☐ 2.   Document ID: US 20040128287 A1
L12: Entry 2 of 2   File: PGPB   Jul 1, 2004

PGPUB-DOCUMENT-NUMBER: 20040128287
PGPUB-FILING-TYPE: new
DOCUMENT-IDENTIFIER: US 20040128287 A1

TITLE: Self tuning database retrieval optimization using regression functions

| Full | Title | Citation | Front | Review | Classification | Date | Reference | Sequences | Attachments | Claims | KWIC | Draw D |

Clear   Generate Collection   Print   Fwd Refs   Bkwd Refs   Generate OACS

Terms	Documents
L9 and (regression same function same (updat$3 or chang$3 or modif$4))	2

Display Format: |- | Change Format

P☮RTAL
USPTO

Search: ⦿ The ACM Digital Library   ○ The Guide

"method estimating query" + "generating sampling" + "regres⟨   **SEARCH**

Feedback   Report a problem   Satisfaction survey

Terms used **method estimating query generating sampling regression function cardinality estimate data mining model**

Found **1** of **184,245**

Sort results by [relevance ▼]

Display results [expanded form ▼]

📖 Save results to a Binder
❓ Search Tips
☐ Open results in a new window

Try an Advanced Search
Try this search in The ACM Guide

Results 1 - 1 of 1

Relevance scale ☐ ▭ ▭ ■ ■

1   Data warehouse design 2: Sense & response service architecture (SARESA): an approach towards a real-time business intelligence solution and its use for a fraud detection application

Tho Manh Nguyen, Josef Schiefer, A. Min Tjoa
November 2005 **Proceedings of the 8th ACM international workshop on Data warehousing and OLAP DOLAP '05**
Publisher: ACM Press
Full text available: 📄 pdf(1.60 MB)   Additional Information: full citation, abstract, references, index terms

The dynamic business environment of many organizations require massive monitoring of their processes in real-time in order to proactively respond to exceptional situations and to take advantage of time-sensitive business opportunities. The ability to sense and interpret events about a changing business environment requires an event-driven IT infrastructure for pwerforming fast and well-informed decisions and putting them into action. However, traditional Business Intelligence (BI) and Data Wareh ...

Keywords: data analysis, event sense & response, real-time business intelligence, real-time data warehousing and OLAP

Results 1 - 1 of 1

Useful downloads: 📄 Adobe Acrobat   ▶ QuickTime   ▶ Windows Media Player   ▶ Real Player

P◉RTAL
USPTO

Search: ◉ The ACM Digital Library   ○ The Guide

"system estimating query" + "generating sampling" + "regress   **SEARCH**

🔺 Feedback   Report a problem   Satisfaction survey

Terms used **system estimating query generating sampling regression function cardinality estimate data mining model**

Found **1** of **184,245**

Sort results by [relevance ▽]

Display results [expanded form ▽]

📗 Save results to a Binder

② Search Tips

☐ Open results in a new window

Try an Advanced Search
Try this search in The ACM Guide

Results 1 - 1 of 1

Relevance scale ☐☐☐■■

1 Data warehouse design 2: Sense & response service architecture (SARESA): an approach towards a real-time business intelligence solution and its use for a fraud detection application   ☐
Tho Manh Nguyen, Josef Schiefer, A. Min Tjoa
November 2005 **Proceedings of the 8th ACM international workshop on Data warehousing and OLAP DOLAP '05**
Publisher: ACM Press
Full text available: 📄 pdf(1.60 MB)   Additional Information: full citation, abstract, references, index terms

The dynamic business environment of many organizations require massive monitoring of their processes in real-time in order to proactively respond to exceptional situations and to take advantage of time-sensitive business opportunities. The ability to sense and interpret events about a changing business environment requires an event-driven IT infrastructure for pwerforming fast and well-informed decisions and putting them into action. However, traditional Business Intelligence (BI) and Data Wareh ...

Keywords: data analysis, event sense & response, real-time business intelligence, real-time data warehousing and OLAP

Results 1 - 1 of 1

Useful downloads: 📄 Adobe Acrobat   ◎ QuickTime   � Windows Media Player   ▷ Real Player

⊙(acm) P⚛RTAL
USPTO

Search: ⊙ The ACM Digital Library   ○ The Guide

"estimating query" + "query condition" + "generating dataset [**SEARCH**]

🔺 Feedback Report a problem Satisfaction survey

Terms used **estimating query query condition generating dataset sampling regression function cardinality estimate data mining model**

Found **2** of **184,245**

Sort results by	relevance ▼
Display results	expanded form ▼

🏷 Save results to a Binder
❓ Search Tips
☐ Open results in a new window

Try an Advanced Search
Try this search in The ACM Guide

Results 1 - 2 of 2

Relevance scale ☐▱▰▰■

1 Query processing: Estimating compilation time of a query optimizer   ☐
Ihab F. Ilyas, Jun Rao, Guy Lohman, Dengfeng Gao, Eileen Lin
June 2003 **Proceedings of the 2003 ACM SIGMOD international conference on Management of data**
Publisher: ACM Press

Full text available: 📄 pdf(292.76 KB)   Additional Information: full citation, abstract, references, citings, index terms

A query optimizer compares alternative plans in its search space to find the best plan for a given query. Depending on the search space and the enumeration algorithm, optimizers vary in their compilation time and the quality of the execution plan they can generate. This paper describes a compilation time estimator that provides a quantified estimate of the optimizer compilation time for a given query. Such an estimator is useful for automatically choosing the right level of optimization in comme ...

2 Research sessions: query progress: Toward a progress indicator for database   ☐
queries
Gang Luo, Jeffrey F. Naughton, Curt J. Ellmann, Michael W. Watzke
June 2004 **Proceedings of the 2004 ACM SIGMOD international conference on Management of data**
Publisher: ACM Press

Full text available: 📄 pdf(228.58 KB)   Additional Information: full citation, abstract, references

Many modern software systems provide progress indicators for long-running tasks. These progress indicators make systems more user-friendly by helping the user quickly estimate how much of the task has been completed and when the task will finish. However, none of the existing commercial RDBMSs provides a non-trival progress indicator for long-running queries. In this paper, we consider the problem of supporting such progress indicators. After discussing the goals and challenges inherent in this ...

Results 1 - 2 of 2

Useful downloads: 📄 Adobe Acrobat   🔵 QuickTime   ▶ Windows Media Player   ▶ Real Player

P⊗RTAL

Search: ⊙ The ACM Digital Library   ○ The Guide

"estimating query" + "correlation query condition" + "generati| **SEARCH**

THE ACM DIGITAL LIBRARY

◆ Feedback Report a problem Satisfaction survey

Terms used **estimating query correlation query condition generating dataset sampling regression function cardinality estimate data mining model**

Found **2** of **184,245**

Sort results by | relevance ▾ |
Display results | expanded form ▾ |

◆ Save results to a Binder
☐ Search Tips
☐ Open results in a new window

Try an **Advanced Search**
Try this search in **The ACM Guide**

Results 1 - 2 of 2

Relevance scale ☐☐☐■■

1 Query processing: Estimating compilation time of a query optimizer   ☐
◆ Ihab F. Ilyas, Jun Rao, Guy Lohman, Dengfeng Gao, Eileen Lin
June 2003 **Proceedings of the 2003 ACM SIGMOD international conference on Management of data**
Publisher: ACM Press

Full text available: 🅐 pdf(292.76 KB)   Additional Information: <u>full citation</u>, <u>abstract</u>, <u>references</u>, <u>citings</u>, <u>index terms</u>

A query optimizer compares alternative plans in its search space to find the best plan for a given query. Depending on the search space and the enumeration algorithm, optimizers vary in their compilation time and the quality of the execution plan they can generate. This paper describes a compilation time estimator that provides a quantified estimate of the optimizer compilation time for a given query. Such an estimator is useful for automatically choosing the right level of optimization in comme ...

2 Research sessions: query progress: Toward a progress indicator for database   ☐
◆ **queries**
Gang Luo, Jeffrey F. Naughton, Curt J. Ellmann, Michael W. Watzke
June 2004 **Proceedings of the 2004 ACM SIGMOD international conference on Management of data**
Publisher: ACM Press

Full text available: 🅐 pdf(228.58 KB)   Additional Information: <u>full citation</u>, <u>abstract</u>, <u>references</u>

Many modern software systems provide progress indicators for long-running tasks. These progress indicators make systems more user-friendly by helping the user quickly estimate how much of the task has been completed and when the task will finish. However, none of the existing commercial RDBMSs provides a non-trival progress indicator for long-running queries. In this paper, we consider the problem of supporting such progress indicators. After discussing the goals and challenges inherent in this ...

Results 1 - 2 of 2

Useful downloads: 🅐 Adobe Acrobat   🅠 QuickTime   ▦ Windows Media Player   ▸ Real Player

P⚙RTAL

USPTO

Search: ⊙ The ACM Digital Library   ○ The Guide

"method estimating query" + "correlation query condition" + " | **SEARCH**

📌 Feedback Report a problem Satisfaction survey

Terms used **method estimating query correlation query condition generating dataset sampling regression function cardinality estimate data mining**

Found **13** of **184,245**

Sort results by | relevance ▼

Display results | expanded form ▼

📚 Save results to a Binder
❓ Search Tips
☐ Open results in a new window

Try an Advanced Search
Try this search in The ACM Guide

Results 1 - 13 of 13

Relevance scale ☐➖➖◼◼

1 Support vector machines: hype or hallelujah?
Kristin P. Bennett, Colin Campbell
December 2000 **ACM SIGKDD Explorations Newsletter**, Volume 2 Issue 2
Publisher: ACM Press
Full text available: 📄 pdf(1.26 MB)   Additional Information: full citation, citings, index terms

☐

Keywords: Support Vector Machines, kernel methods, statistical learning theory

2 Industrial session: new data types and algorithms: SVM in oracle database 10g: removing the barriers to widespread adoption of support vector machines
Boriana L. Milenova, Joseph S. Yarmus, Marcos M. Campos
August 2005 **Proceedings of the 31st international conference on Very large data bases VLDB '05**
Publisher: VLDB Endowment
Full text available: 📄 pdf(190.75 KB)   Additional Information: full citation, abstract, references, index terms

☐

Contemporary commercial databases are placing an increased emphasis on analytic capabilities. Data mining technology has become crucial in enabling the analysis of large volumes of data. Modern data mining techniques have been shown to have high accuracy and good generalization to novel data. However, achieving results of good quality often requires high levels of user expertise. Support Vector Machines (SVM) is a powerful state-of-the-art data mining algorithm that can address problems not amen ...

3 Research track: Online novelty detection on temporal sequences
Junshui Ma, Simon Perkins
August 2003 **Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining**
Publisher: ACM Press
Full text available: 📄 pdf(800.25 KB)   Additional Information: full citation, abstract, references, citings, index terms

☐

In this paper, we present a new framework for online novelty detection on temporal sequences. This framework include a mechanism for associating each detection result with a confidence value. Based on this framework, we develop a concrete online detection

algorithm, by modeling the temporal sequence using an online support vector regression algorithm. Experiments on both synthetic and real world data are performed to demonstrate the promising performance of our proposed detection algorithm.

Keywords: anomaly detection, novelty detection, online algorithm, support vector regression

4 Learning classifiers: Learning block importance models for web pages ☐

Ruihua Song, Haifeng Liu, Ji-Rong Wen, Wei-Ying Ma

May 2004 **Proceedings of the 13th international conference on World Wide Web**

Publisher: ACM Press

Full text available: 📄 pdf(1.23 MB)   Additional Information: <u>full citation</u>, <u>abstract</u>, <u>references</u>, <u>citings</u>, <u>index terms</u>

Previous work shows that a web page can be partitioned into multiple segments or blocks, and often the importance of those blocks in a page is not equivalent. Also, it has been proven that differentiating noisy or unimportant blocks from pages can facilitate web mining, search and accessibility. However, no uniform approach and model has been presented to measure the importance of different segments in web pages. Through a user study, we found that people do have a consistent view about the impo ...

Keywords: block importance model, classification, page segmentation, web mining

5 Instance-based attribute identification in database integration ☐

Cecil Eng H. Chua, Roger H. L. Chiang, Ee-Peng Lim

October 2003 **The VLDB Journal — The International Journal on Very Large Data Bases**, Volume 12 Issue 3

Publisher: Springer-Verlag New York, Inc.

Full text available: 📄 pdf(220.13 KB)   Additional Information: <u>full citation</u>, <u>abstract</u>, <u>index terms</u>

Abstract.Most research on attribute identification in database integration has focused on integrating attributes using schema and summary information derived from the attribute values. No research has attempted to fully explore the use of attribute values to perform attribute identification. We propose an attribute identification method that employs schema and summary instance information as well as properties of attributes derived from their instances. Unlike other attribute identification meth ...

Keywords: Attribute identification, Database integration, Measures of association

6 Database technology for novel applications: MauveDB: supporting model-based user ☐ views in database systems

Amol Deshpande, Samuel Madden

June 2006 **Proceedings of the 2006 ACM SIGMOD international conference on Management of data SIGMOD '06**

Publisher: ACM Press

Full text available: 📄 pdf(589.31 KB)   Additional Information: <u>full citation</u>, <u>abstract</u>, <u>references</u>, <u>index terms</u>

Real-world data --- especially when generated by distributed measurement infrastructures such as sensor networks --- tends to be incomplete, imprecise, and erroneous, making it impossible to present it to users or feed it directly into applications. The traditional approach to dealing with this problem is to first process the data using statistical or probabilistic *models* that can provide more robust interpretations of the data. Current database systems, however, do not provide adequate s ...

Keywords: query processing, regression, sensor networks, statistical models, uncertain data, views

7 Learning important models for web page blocks based on layout and content analysis ☐
Ruihua Song, Haifeng Liu, Ji-Rong Wen, Wei-Ying Ma
December 2004 **ACM SIGKDD Explorations Newsletter**, Volume 6 Issue 2
Publisher: ACM Press
Full text available: 📄 pdf(601.78 KB)   Additional Information: full citation, abstract, references

Previous work shows that a web page can be partitioned into multiple segments or blocks, and often the importance of those blocks in a page is not equivalent. It has also been proven that differentiating noisy and unimportant blocks from pages can facilitate web mining, search and accessibility. However, no uniform approach and model has been presented to measure the importance of different blocks in a web page. Through a user study, we found that people do have a consistent view about the impor ...

Keywords: block importance model, classification, page segmentation, web mining

8 Adaptive performance prediction for distributed data-intensive applications ☐
Marcio Faerman, Alan Su, Richard Wolski, Francine Berman
January 1999 **Proceedings of the 1999 ACM/IEEE conference on Supercomputing (CDROM)**
Publisher: ACM Press
Full text available: 📄 pdf(292.25 KB)   Additional Information: full citation, references, citings, index terms

9 Articles on microarray data mining: Loss-based estimation with cross-validation: applications to microarray data analysis ☐
Sandrine Dudoit, Mark J. van der Laan, Sündüz Keleş, Annette M. Molinaro, Sandra E. Sinisi, Siew Leng Teng
December 2003 **ACM SIGKDD Explorations Newsletter**, Volume 5 Issue 2
Publisher: ACM Press
Full text available: 📄 pdf(325.39 KB)   Additional Information: full citation, abstract, references, citings

Current statistical inference problems in genomic data analysis involve parameter estimation for high-dimensional multivariate distributions, with typically unknown and intricate correlation patterns among variables. Addressing these inference questions satisfactorily requires: (i) an intensive and thorough search of the parameter space to generate good candidate estimators; (ii) an approach for selecting an optimal estimator among these candidates; and (iii) a method for reliably assessing the ...

Keywords: censored data, classification, comparative genomic hybridization, cross-validation, density estimation, estimation, loss function, microarray, model selection, multivariate outcome, prediction, regression trees, risk, survival analysis, variable selection

10 A Modified Finite Newton Method for Fast Solution of Large Scale Linear SVMs ☐
S. Sathiya Keerthi, Dennis DeCoste
September 2005 **The Journal of Machine Learning Research**, Volume 6
Publisher: MIT Press
Full text available: 📄 pdf(157.28 KB)   Additional Information: full citation, abstract

This paper develops a fast method for solving linear SVMs with L_2 loss function that is suited for large scale data mining tasks such as text classification. This is done by modifying the finite Newton method of Mangasarian in several ways. Experiments indicate that the method is much faster than decomposition methods such as SVM^{light}, SMO and BSVM (e.g., 4-100 fold), especially when the number of examples is large. The paper also suggests ways of extending the metho ...

11 A Modified Finite Newton Method for Fast Solution of Large Scale Linear SVMs

S. Sathiya Keerthi, Dennis DeCoste
September 2005 **The Journal of Machine Learning Research**, Volume 6
Publisher: MIT Press
Full text available: pdf(195.71 KB)   Additional Information: full citation, abstract, citings

This paper develops a fast method for solving linear SVMs with L_2 loss function that is suited for large scale data mining tasks such as text classification. This is done by modifying the finite Newton method of Mangasarian in several ways. Experiments indicate that the method is much faster than decomposition methods such as SVM^{light}, SMO and BSVM (e.g., 4-100 fold), especially when the number of examples is large. The paper also suggests ways of extending the metho ...

12 Statistical methods I: MARK: a boosting algorithm for heterogeneous kernel models

Kristin P. Bennett, Michinari Momma, Mark J. Embrechts
July 2002 **Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining**
Publisher: ACM Press

Full text available: pdf(774.00 KB)   Additional Information: full citation, abstract, references, citings, index terms

Support Vector Machines and other kernel methods have proven to be very effective for nonlinear inference. Practical issues are how to select the type of kernel including any parameters and how to deal with the computational issues caused by the fact that the kernel matrix grows quadratically with the data. Inspired by ensemble and boosting methods like MART, we propose the Multiple Additive Regression Kernels (MARK) algorithm to address these issues. MARK considers a large (potentially infinite ...

13 Research papers: optimization: Proactive re-optimization

Shivnath Babu, Pedro Bizarro, David DeWitt
June 2005 **Proceedings of the 2005 ACM SIGMOD international conference on Management of data**
Publisher: ACM Press
Full text available: pdf(419.53 KB)   Additional Information: full citation, abstract, references

Traditional query optimizers rely on the accuracy of estimated statistics to choose good execution plans. This design often leads to suboptimal plan choices for complex queries, since errors in estimates for intermediate subexpressions grow exponentially in the presence of skewed and correlated data distributions. Reoptimization is a promising technique to cope with such mistakes. Current re-optimizers first use a traditional optimizer to pick a plan, and then react to estimation errors and resu ...

Results 1 - 13 of 13

Useful downloads: Adobe Acrobat   QuickTime   Windows Media Player   Real Player

P⊛RTAL

USPTO

Search: ⊚ The ACM Digital Library   ○ The Guide

"system estimating query" + "correlation query condition" + "c   **SEARCH**

█ Feedback Report a problem Satisfaction survey

Terms used **system estimating query correlation query condition generating dataset sampling regression function cardinality estimate data mining**

Found **13** of **184,245**

Sort results by [relevance ▽]

Display results [expanded form ▽]

◆ Save results to a Binder
? Search Tips
☐ Open results in a new window

Try an Advanced Search
Try this search in The ACM Guide

Results 1 - 13 of 13

Relevance scale ☐☐☐■■

1 Support vector machines: hype or hallelujah?   ☐
Kristin P. Bennett, Colin Campbell
December 2000 **ACM SIGKDD Explorations Newsletter,** Volume 2 Issue 2
Publisher: ACM Press
Full text available: 📄 pdf(1.26 MB)   Additional Information: full citation, citings, index terms

Keywords: Support Vector Machines, kernel methods, statistical learning theory

2 Industrial session: new data types and algorithms: SVM in oracle database 10g: ☐
removing the barriers to widespread adoption of support vector machines
Boriana L. Milenova, Joseph S. Yarmus, Marcos M. Campos
August 2005 **Proceedings of the 31st international conference on Very large data**
      bases VLDB '05
Publisher: VLDB Endowment
Full text available: 📄 pdf(190.75 KB)   Additional Information: full citation, abstract, references, index terms

Contemporary commercial databases are placing an increased emphasis on analytic
capabilities. Data mining technology has become crucial in enabling the analysis of large
volumes of data. Modern data mining techniques have been shown to have high accuracy
and good generalization to novel data. However, achieving results of good quality often
requires high levels of user expertise. Support Vector Machines (SVM) is a powerful state-
of-the-art data mining algorithm that can address problems not amen ...

3 Research track: Online novelty detection on temporal sequences   ☐
Junshui Ma, Simon Perkins
August 2003 **Proceedings of the ninth ACM SIGKDD international conference on**
      Knowledge discovery and data mining
Publisher: ACM Press
Full text available: 📄 pdf(800.25 KB)   Additional Information: full citation, abstract, references, citings, index
                                                                                    terms

In this paper, we present a new framework for online novelty detection on temporal
sequences. This framework include a mechanism for associating each detection result with
a confidence value. Based on this framework, we develop a concrete online detection

algorithm, by modeling the temporal sequence using an online support vector regression algorithm. Experiments on both synthetic and real world data are performed to demonstrate the promising performance of our proposed detection algorithm.

Keywords: anomaly detection, novelty detection, online algorithm, support vector regression

4 Learning classifiers: Learning block importance models for web pages

Ruihua Song, Haifeng Liu, Ji-Rong Wen, Wei-Ying Ma

May 2004 **Proceedings of the 13th international conference on World Wide Web**

Publisher: ACM Press

Full text available: pdf(1.23 MB)   Additional Information: full citation, abstract, references, citings, index terms

Previous work shows that a web page can be partitioned into multiple segments or blocks, and often the importance of those blocks in a page is not equivalent. Also, it has been proven that differentiating noisy or unimportant blocks from pages can facilitate web mining, search and accessibility. However, no uniform approach and model has been presented to measure the importance of different segments in web pages. Through a user study, we found that people do have a consistent view about the impo ...

Keywords: block importance model, classification, page segmentation, web mining

5 Instance-based attribute identification in database integration

Cecil Eng H. Chua, Roger H. L. Chiang, Ee-Peng Lim

October 2003 **The VLDB Journal — The International Journal on Very Large Data Bases**, Volume 12 Issue 3

Publisher: Springer-Verlag New York, Inc.

Full text available: pdf(220.13 KB)   Additional Information: full citation, abstract, index terms

Abstract.Most research on attribute identification in database integration has focused on integrating attributes using schema and summary information derived from the attribute values. No research has attempted to fully explore the use of attribute values to perform attribute identification. We propose an attribute identification method that employs schema and summary instance information as well as properties of attributes derived from their instances. Unlike other attribute identification meth ...

Keywords: Attribute identification, Database integration, Measures of association

6 Database technology for novel applications: MauveDB: supporting model-based user views in database systems

Amol Deshpande, Samuel Madden

June 2006 **Proceedings of the 2006 ACM SIGMOD international conference on Management of data SIGMOD '06**

Publisher: ACM Press

Full text available: pdf(589.31 KB)   Additional Information: full citation, abstract, references, index terms

Real-world data --- especially when generated by distributed measurement infrastructures such as sensor networks --- tends to be incomplete, imprecise, and erroneous, making it impossible to present it to users or feed it directly into applications. The traditional approach to dealing with this problem is to first process the data using statistical or probabilistic *models* that can provide more robust interpretations of the data. Current database systems, however, do not provide adequate s ...

Keywords: query processing, regression, sensor networks, statistical models, uncertain data, views

7   Learning important models for web page blocks based on layout and content analysis ☐
Ruihua Song, Haifeng Liu, Ji-Rong Wen, Wei-Ying Ma
December 2004 **ACM SIGKDD Explorations Newsletter**, Volume 6 Issue 2
Publisher: ACM Press
Full text available: 📄 pdf(601.78 KB)   Additional Information: full citation, abstract, references

Previous work shows that a web page can be partitioned into multiple segments or blocks, and often the importance of those blocks in a page is not equivalent. It has also been proven that differentiating noisy and unimportant blocks from pages can facilitate web mining, search and accessibility. However, no uniform approach and model has been presented to measure the importance of different blocks in a web page. Through a user study, we found that people do have a consistent view about the impor ...

Keywords: block importance model, classification, page segmentation, web mining

8   Adaptive performance prediction for distributed data-intensive applications ☐
Marcio Faerman, Alan Su, Richard Wolski, Francine Berman
January 1999 **Proceedings of the 1999 ACM/IEEE conference on Supercomputing (CDROM)**
Publisher: ACM Press
Full text available: 📄 pdf(292.25 KB)   Additional Information: full citation, references, citings, index terms

9   Articles on microarray data mining: Loss-based estimation with cross-validation: applications to microarray data analysis ☐
Sandrine Dudoit, Mark J. van der Laan, Sündüz Keleş, Annette M. Molinaro, Sandra E. Sinisi, Siew Leng Teng
December 2003 **ACM SIGKDD Explorations Newsletter**, Volume 5 Issue 2
Publisher: ACM Press
Full text available: 📄 pdf(325.39 KB)   Additional Information: full citation, abstract, references, citings

Current statistical inference problems in genomic data analysis involve parameter estimation for high-dimensional multivariate distributions, with typically unknown and intricate correlation patterns among variables. Addressing these inference questions satisfactorily requires: (i) an intensive and thorough search of the parameter space to generate good candidate estimators; (ii) an approach for selecting an optimal estimator among these candidates; and (iii) a method for reliably assessing the ...

Keywords: censored data, classification, comparative genomic hybridization, cross-validation, density estimation, estimation, loss function, microarray, model selection, multivariate outcome, prediction, regression trees, risk, survival analysis, variable selection

10   A Modified Finite Newton Method for Fast Solution of Large Scale Linear SVMs ☐
S. Sathiya Keerthi, Dennis DeCoste
September 2005 **The Journal of Machine Learning Research**, Volume 6
Publisher: MIT Press
Full text available: 📄 pdf(157.28 KB)   Additional Information: full citation, abstract

This paper develops a fast method for solving linear SVMs with L_2 loss function that is suited for large scale data mining tasks such as text classification. This is done by modifying the finite Newton method of Mangasarian in several ways. Experiments indicate that the method is much faster than decomposition methods such as SVM^{light}, SMO and BSVM (e.g., 4-100 fold), especially when the number of examples is large. The paper also suggests ways of extending the metho ...

11 [A Modified Finite Newton Method for Fast Solution of Large Scale Linear SVMs](#)

S. Sathiya Keerthi, Dennis DeCoste

September 2005 **The Journal of Machine Learning Research**, Volume 6

Publisher: MIT Press

Full text available: pdf(195.71 KB)   Additional Information: full citation, abstract, citings

This paper develops a fast method for solving linear SVMs with L_2 loss function that is suited for large scale data mining tasks such as text classification. This is done by modifying the finite Newton method of Mangasarian in several ways. Experiments indicate that the method is much faster than decomposition methods such as SVM^{light}, SMO and BSVM (e.g., 4-100 fold), especially when the number of examples is large. The paper also suggests ways of extending the metho ...

12 [Statistical methods I: MARK: a boosting algorithm for heterogeneous kernel models](#)

Kristin P. Bennett, Michinari Momma, Mark J. Embrechts

July 2002 **Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining**

Publisher: ACM Press

Full text available: pdf(774.00 KB)   Additional Information: full citation, abstract, references, citings, index terms

Support Vector Machines and other kernel methods have proven to be very effective for nonlinear inference. Practical issues are how to select the type of kernel including any parameters and how to deal with the computational issues caused by the fact that the kernel matrix grows quadratically with the data. Inspired by ensemble and boosting methods like MART, we propose the Multiple Additive Regression Kernels (MARK) algorithm to address these issues. MARK considers a large (potentially infinite ...

13 [Research papers: optimization: Proactive re-optimization](#)

Shivnath Babu, Pedro Bizarro, David DeWitt

June 2005 **Proceedings of the 2005 ACM SIGMOD international conference on Management of data**

Publisher: ACM Press

Full text available: pdf(419.53 KB)   Additional Information: full citation, abstract, references

Traditional query optimizers rely on the accuracy of estimated statistics to choose good execution plans. This design often leads to suboptimal plan choices for complex queries, since errors in estimates for intermediate subexpressions grow exponentially in the presence of skewed and correlated data distributions. Reoptimization is a promising technique to cope with such mistakes. Current re-optimizers first use a traditional optimizer to pick a plan, and then react to estimation errors and resu ...

Results 1 - 13 of 13

IEEE *Xplore*®
RELEASE 2.1

Welcome United States Patent and Trademark Office

▓░▓Search Results   **BROWSE   SEARCH   IEEE XPLORE GUIDE**

Results for "(((estimat* <paragraph> (query* <near> condition)) <and> cardinality)<in>met..."   ✉ e-mail
Your search matched **0** documents.
A maximum of **100** results are displayed, **25** to a page, sorted by **Relevance** in **Descending** order.

» Search Options

View Session History   **Modify Search**

New Search | (((estimat* <paragraph> (query* <near> condition)) <and> cardinality)<in>metadata) | **Search**

☐ Check to search only within this results set

» Key   **Display Format:** ⦿ Citation ○ Citation & Abstract

IEEE JNL   IEEE Journal or
              Magazine

IEE JNL   IEE Journal or Magazine

IEEE CNF   IEEE Conference   **No results were found.**
              Proceeding
                                    Please edit your search criteria and try again. Refer to the Help pages if you need assistan
IEE CNF   IEE Conference   search.
              Proceeding

IEEE STD   IEEE Standard

Indexed by
▓ Inspec®

IEEE Xplore ®
RELEASE 2.1

Welcome United States Patent and Trademark Office

Search Results   BROWSE   SEARCH   IEEE XPLORE GUIDE

Results for "(((estimat* <paragraph> (query* <near> condition)) <and> cardinalities)<in>m..."   ✉ e-mail
Your search matched **0** documents.
A maximum of **100** results are displayed, **25** to a page, sorted by **Relevance** in **Descending** order.

» Search Options

View Session History

New Search

» Key

IEEE JNL   IEEE Journal or
Magazine

IEE JNL   IEE Journal or Magazine

IEEE CNF   IEEE Conference
Proceeding

IEE CNF   IEE Conference
Proceeding

IEEE STD   IEEE Standard

Modify Search

(((estimat* <paragraph> (query* <near> condition)) <and> cardinalities)<in>metadata   **Search**

☐ Check to search only within this results set

Display Format: ⦿ Citation   ○ Citation & Abstract

No results were found.

Please edit your search criteria and try again. Refer to the Help pages if you need assistan
search.

Help   Contact Us   Privacy & :

Indexed by
Inspec®

IEEE *Xplore* ⑫
RELEASE 2.1

Welcome United States Patent and Trademark Office

⊡Search Results

BROWSE   SEARCH   IEEE XPLORE GUIDE

Results for "(((estimat* <paragraph> query*) <and> (regression <near> function))<in>metad..."   ☑e-mail
Your search matched **4** of **1392165** documents.
A maximum of **100** results are displayed, **25** to a page, sorted by **Relevance** in **Descending** order.

» Search Options

View Session History

New Search

» Key

IEEE JNL	IEEE Journal or Magazine
IEE JNL	IEE Journal or Magazine
IEEE CNF	IEEE Conference Proceeding
IEE CNF	IEE Conference Proceeding
IEEE STD	IEEE Standard

Modify Search

(((estimat* <paragraph> query*) <and> (regression <near> function))<in>metadata) [Search]

☐ Check to search only within this results set

Display Format: ⦿ Citation ○ Citation & Abstract

[view selected items]   **Select All Deselect All**

☐ 1. **Image retrieval based on similarity learning**
El-Naqa, I.; Wernick, M.N.; Yongyi Yang; Galatsanos, N.P.;
Image Processing, 2000. Proceedings. 2000 International Conference on
Volume 3, 10-13 Sept. 2000 Page(s):722 - 725 vol.3
Digital Object Identifier 10.1109/ICIP.2000.899556

AbstractPlus | Full Text: PDF(312 KB)   **IEEE CNF**
Rights and Permissions

☐ 2. **Gaussian process regression: active data selection and test point rejecti**
Sambu Seo; Wallat, M.; Graepel, T.; Obermayer, K.;
Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS In
Conference on
Volume 3, 24-27 July 2000 Page(s):241 - 246 vol.3
Digital Object Identifier 10.1109/IJCNN.2000.861310

AbstractPlus | Full Text: PDF(412 KB)   **IEEE CNF**
Rights and Permissions

☐ 3. **Stochastic simulations of rejected World Wide Web pages**
Meghabghab, G.;
Modeling, Analysis and Simulation of Computer and Telecommunication Syste
Proceedings. 8th International Symposium on
29 Aug.-1 Sept. 2000 Page(s):483 - 491
Digital Object Identifier 10.1109/MASCOT.2000.876575

AbstractPlus | Full Text: PDF(1064 KB)   **IEEE CNF**
Rights and Permissions

☐ 4. **Approximate query answering in numerical databases**
Hachem, N.; Chenye Bao; Taylor, S.;
Scientific and Statistical Database Systems, 1996. Proceedings., Eighth Intern:
Conference on
18-20 June 1996 Page(s):63 - 73
Digital Object Identifier 10.1109/SSDM.1996.505916

AbstractPlus | Full Text: PDF(880 KB)   **IEEE CNF**
Rights and Permissions

Help   Contact Us   Privacy & :

Welcome United States Patent and Trademark Office

CIISearch Results   BROWSE   SEARCH   IEEE XPLORE GUIDE

Results for "(((estimat* <near> selective <near> query*) <and> (data <near> min* <near..."   ☑e-mail
Your search matched **1** of **1392165** documents.
A maximum of **100** results are displayed, **25** to a page, sorted by **Relevance** in **Descending** order.

» Search Options

View Session History

New Search

Modify Search

(((estimat* <near> selective <near> query*) <and> (data <near> min* <near> model))   **Search**

☐ Check to search only within this results set

Display Format: ⦿ Citation   ○ Citation & Abstract

» Key

IEEE JNL	IEEE Journal or Magazine
IEE JNL	IEE Journal or Magazine
IEEE CNF	IEEE Conference Proceeding
IEE CNF	IEE Conference Proceeding
IEEE STD	IEEE Standard

view selected items   **Select All Deselect All**

☐   **1. Efficient XML query processing in mediators**
Liang Huai Yang; Shiwei Tang; Dongqing Yang; Lijun Chen;
Database and Expert Systems Applications, 2001. Proceedings. 12th Internatii
3-7 Sept. 2001 Page(s):27 - 31
Digital Object Identifier 10.1109/DEXA.2001.953037

AbstractPlus | Full Text: PDF(336 KB)   IEEE CNF
Rights and Permissions

Help   Contact Us   Privacy & :
© Copyright 2006 IEEE –

Indexed by
⌷ Inspec°